**TCP FIN\_WAIT\_2探究(1)**

tcp close简单来说只是四次挥手，但在四次挥手过程中，如果其中一端断电、系统崩溃，可能会引发另一端端口长时间释放不了而占用系统资源，下面我会针对tcp FIN\_WAIT\_2做一些说明:

与fin\_wait\_2超时时间相关的参数有struct tcp\_sock的linger2变量和sysctl\_tcp\_fin\_timeout变量，linger2变量声明如下所示:

struct tcp\_sock {

int linger2;

};

linger2通过setsockopt的SOL\_TCP，TCP\_LINGER2设置，单位是jiffies；setsockopt设置TCP\_LINGER2传入的val参数单位是秒，通过HZ\*val转换成jiffies赋值给linger2;在设置linger2时，如果传入的val小于0，则置linger2为-1；如果传入的val值比tcp\_fin\_timeout大，则置linger2为0；如果上述两点都不满足，则置linger2为val\*HZ；

sysctl\_tcp\_fin\_timeout在/proc/sys/net/ipv4下有对应的可修改项:tcp\_fin\_timeout,tcp\_fin\_timeout通过proc接口读的是以秒为单位，在内核对应的sysctl\_tcp\_fin\_timeout变量以jiffies为单位，在net/ipv4/sysctl\_net\_ipv4.c文件中可见sysctl\_tcp\_fin\_timeout的proc\_handler是proc\_dointvec\_jiffies，所以proc接口会对其有个转换；

tcp在fin\_wait\_2状态超过linger2对应的时间并且tcp是SOCK\_DEAD状态，会给对端发送reset并且释放当前端口；用户可以通过setsockopt和修改tcp\_fin\_timeout来修改linger2的时间;

struct tcp\_sock的linger2成员在初始化时随struct tcp\_sock kzmalloc为0；sysclt\_tcp\_fin\_timeout在net/ipv4/tcp\_ipv4.c tcp\_sk\_init中初始化，一般和timewait时间一样为60s；

在本地tcp发送fin后，本地tcp进入fin\_wait\_1状态，在本地tcp收到发出去的fin的ack后，本地tcp进入fin\_wait\_2状态，在进入fin\_wait\_2状态后，tcp可能会出现如下几种情况:

1. 如果当前连接的linger2小于0，则给对端发送reset、释放本地端口并增加TCPABORTDATA mibs
2. 如果linger2或sysctl\_tcp\_fin\_timeout（linger2优先）大于TCP\_TIMEWAIT\_LEN(60s),则启动keepalive超时定时器，超时时间设置为fin\_wait\_2 超时时间减去TCP\_TIMEWAIT\_LEN时间；如果keepalive定时器超时，struct sock的sk\_flags成员 SOCK\_DEAD置位，则又进入timewait超时，超时时间也是fin\_wait\_2超时减去TCP\_TIMEWAIT\_LEN时间，timewait的tw\_substate被置为TCP\_FIN\_WAIT\_2；如果timewait超时，则会释放该端口；
3. 如果linger2或sysctl\_tcp\_fin\_timeout(linger2优先)大于TCP\_TIMEWAIT\_LEN(60s),则启动keepalive超时定时器，超时时间设置为fin\_wait\_2超时时间减去TCP\_TIMEWAIT\_LEN时间；如果keepalive定时器超时前收到对端报文，则会通过如下的函数调用栈进入sk\_data\_queue:

tcp\_v4\_rcv->

tcp\_v4\_do\_rcv->

tcp\_rcv\_state\_process->

tcp\_data\_queue

在tcp\_data\_queue中如果检测到fin被置位，则又进入timewait状态超时。

1. 如果linger2小于等于TCP\_TIMEWAIT\_LEN(60s)，则直接进入timewait超时，如果timewait超时时间到，则释放该端口
2. 如果linger2小于等于TCP\_TIMEWAIT\_LEN(60s)，则直接进入timewait超时，如果timewait超时前收到报文，则进入如下流程:

tcp\_v4\_rcv->

tcp\_timewait\_state\_process

在tcp\_timewait\_state\_process中如果检测到收到fin报文，则进入真正的timewait阶段